CLAIMS

What is claimed is:

1. A session initiation protocol (SIP) signaling router comprising:

   (a) a plurality of cluster nodes for performing at least one session initiation protocol function, each cluster node storing a local database containing SIP location information; and

   (b) a location server coupled to the cluster nodes for maintaining a database of SIP location information and for automatically replicating the database of SIP location information to each of the cluster nodes in real time in response to receiving updates to the SIP location information.

2. The SIP signaling router of claim 1 wherein each of the cluster nodes comprises a SIP proxy server.

3. The SIP signaling router of claim 1 wherein each of the cluster nodes comprises a SIP redirect server.

4. The SIP signaling router of claim 1 wherein each of the cluster nodes comprises a SIP proxy server and a SIP redirect server.

5. The SIP signaling router of claim 1 wherein the location server is adapted to replicate the database of SIP location information to each of the cluster nodes using a reliable multicast transport protocol (RMTP).

6. The SIP signaling router of claim 1 comprising first and second layer 2 switches coupled to each of the cluster nodes.

7. The SIP signaling router of claim 6 wherein each of the cluster nodes include first and second network interfaces and the first layer 2 switch is

coupled to the first network interface of each of the cluster nodes and the second layer 2 switch is coupled to the second network interface of each of the cluster nodes.

8. The SIP signaling router of claim 7 wherein at least one of the first and second layer 2 switches is configured to periodically ping each of the cluster nodes to determine sub-application level protocol stack operational status of the cluster nodes.

9. The SIP signaling router of claim 7 wherein the first layer 2 switch is adapted to periodically send health check messages to each of the cluster nodes to determine application-level operational status.

10. The SIP signaling router of claim 9 wherein the first layer 2 switch is adapted to determine the operational status based on the response time of each of the cluster nodes to the health check messages.

11. The SIP signaling router of claim 6 wherein the first and second layer 2 switches are redundantly connected to each of the cluster nodes.

12. The SIP signaling router of claim 11 wherein the first and second layer 2 switches are adapted to dynamically reroute SIP signaling traffic around congested or failed signaling links using a link aggregation control protocol.

13. The SIP signaling router of claim 1 wherein the location server includes:

(a) a provisioning database task for provisioning the SIP location information in the database;

(b) a database provisioning log for tracking changes to the database; and

(c) a network provisioning task for detecting updates to the database based on the database provisioning log and for distributing the updates to the cluster nodes in real time in response to detecting the updates.

14. The SIP signaling router of claim 13 wherein the network provisioning task is adapted to multicast the updates to the cluster nodes.

15. The SIP signaling router of claim 13 wherein the network provisioning task is adapted to multicast the updates to the cluster nodes using the reliable multicast transport protocol.

16. The SIP signaling router of claim 1 wherein the cluster nodes and the location server each comprise stand alone computers or workstations.

17. The SIP signaling router of claim 1 further comprising an interprocessor message transport bus for carrying message between the cluster nodes and the location server, wherein the cluster nodes and the location server each comprise a printed circuit board connected to the interprocessor message transport bus.

18. A scalable, reliable, session initiation protocol (SIP) signaling router, comprising:

(a) a plurality of cluster nodes, each of the cluster nodes adapted to perform at least one SIP protocol function;

(b) a primary Ethernet switch coupled to each of the cluster nodes; and

(c) a secondary Ethernet switch coupled to each of the cluster nodes, wherein the primary Ethernet switch is adapted to send PING and health check messages to each of the cluster nodes to determine

the operational status of each of the cluster nodes and perform a load sharing operation among the cluster nodes based on the operational status.

19. The SIP signaling router of claim 18 wherein the primary Ethernet switch is adapted to determine the operational status of each of the cluster nodes by measuring the response time of each of the cluster nodes to the PING and health check messages.

20. The SIP signaling router of claim 18 wherein the primary Ethernet switch maintains a connection tuple table containing information regarding active connections with each of the cluster nodes.

21. The SIP signaling router of claim 20 wherein the primary Ethernet switch determines the operational status based on the number of active connections for each cluster node maintained in the connection tuple table.

22. The SIP signaling router of claim 20 wherein the primary Ethernet switch replicates the connection tuple table to the secondary Ethernet switch.

23. The SIP signaling router of claim 22 wherein, in response to failure of the primary Ethernet switch, the secondary Ethernet switches to function as the primary Ethernet switch.

24. A method for routing session initiation protocol (SIP) signaling messages, the method comprising:

(a) replicating a database of SIP location information from a location server to a plurality of cluster nodes, each cluster node performing a SIP protocol function;

(b) receiving SIP signaling messages at the cluster nodes for requesting SIP protocol services; and

(c) determining SIP location information for the SIP signaling messages in real time using the local databases at the cluster nodes without querying the location server.

25. The method of claim 24 wherein replicating a database of SIP location information to the cluster nodes includes notifying the cluster nodes of incremental changes in the database level at the location server and incrementally updating the location databases maintained by the cluster nodes as changes are made to the database maintained by the location server.

26. The method of claim 24 wherein replicating a database of SIP location information to the cluster nodes includes reloading the database on each of the cluster nodes when a database level difference between the cluster nodes and the location server exceeds a predetermined threshold.

27. The method of claim 24 wherein replicating a database of SIP location information to the cluster nodes includes forwarding database updates to the cluster nodes using the reliable multicast transport protocol.

28. The method of claim 24 wherein the SIP signaling messages include SIP INVITE messages.

29. A method for load sharing among cluster nodes providing session initiation protocol (SIP) services, the method comprising:

(a) providing a plurality of cluster nodes for performing SIP services;

(b) periodically sending messages to each of the cluster nodes to monitor the operational status of each of the cluster nodes;

(c) maintaining a connection tuple table containing active connection information for each of the cluster nodes; and

(d) load sharing between the cluster nodes based on the operational status and the connection tuple table.

30. The method of claim 29 wherein providing a plurality of cluster nodes includes providing a plurality of cluster nodes for performing SIP proxy services.

31. The method of claim 29 wherein providing a plurality of cluster nodes includes providing a plurality of cluster nodes for performing SIP redirect services.

32. The method of claim 29 wherein providing a plurality of cluster nodes includes providing a plurality of cluster nodes for performing SIP proxy and redirect services.

33. The method of claim 29 wherein periodically sending messages to the cluster nodes includes periodically pinging the cluster nodes to determine lower-level protocol stack operational status of each of the cluster nodes.

34. The method of claim 29 wherein periodically sending messages to the cluster nodes includes periodically sending health check messages to the cluster nodes to determine application-level protocol stack operational status of each of the cluster nodes.

35. The method of claim 29 wherein periodically sending messages to the cluster nodes includes periodically pinging and sending health check messages to the cluster nodes to determine lower-level protocol stack and application level operational status of each of the cluster nodes.

36. The method of claim 29 wherein maintaining a connection tuple table for each of the cluster nodes includes maintaining a table containing destination IP address, source IP address, destination port number, originating port number, and MAC address of the cluster node for each connection.

37. The method of claim 29 wherein load sharing based on the operational status and the connection tuple table includes load sharing based each cluster node's response time to the messages and the number of active connections for each node stored in the connection tuple table.